Calhoun: The NPS Institutional Archive

DSpace Repository

Theses and Dissertations | 1. Thesis and Dissertation Collection, all items

1970

The usefulness of task analysis in the evaluation of military training.

Stewart, Joseph David

Monterey, California ; Naval Postgraduate School

http://hdl.handle.net/10945/15207

Downloaded from NPS Archive: Calhoun

http://www.nps.edu/library

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

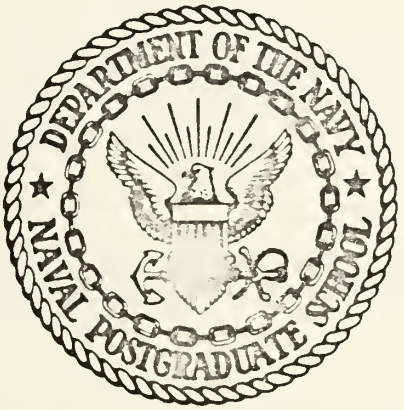
THE USEFULNESS OF TASK ANALYSIS IN THE EVALUATION OF MILITARY TRAINING

by

Joseph David Stewart

KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93943-5002

United States
Naval Postgraduate School



THESIS

THE USEFULNESS OF TASK ANALYSIS IN THE EVALUATION OF MILITARY TRAINING

by

Joseph David Stewart

September 1970

This document has been approved for public release and sale; its distribution is unlimited.

T135116

The Usefulness of Task Analysis in the Evaluation of Military Training

by

Joseph David Stewart
Captain, United States Marine Corps
B.S., United States Naval Academy, 1964

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1970

RARY
L POSTGRADUATE SCHOOL
TEREY, CALIF. 93940

ABSTRACT

Determining the proper emphasis of curriculum contents as well as judging the value or worth of training programs has become an important problem. The purpose of this paper is to demonstrate the usefulness of task analysis in measuring the effectiveness of training courses based on the extent to which curriculum contents are job oriented. In this regard, parametric and nonparametric statistical procedures are discussed as well as a matrix method of evaluation. A general methodology to include the operational significance of the data is also included.

Additionally, the results of a small scale experiment are presented. This experiment was conducted to determine the most valid questionnaire associated data collection method.

TABLE OF CONTENTS

I. INTRODUCTION ---------- 9

A. SCOPE ---------- 11

B. ASSUMPTIONS ---------- 14

C. METHODOLOGY ---------- 19

D. DATA COLLECTION METHODS ---------- 21

II. STATISTICAL TESTING IN GENERAL ---------- 23

III. NONPARAMETRIC EVALUATION OF TRAINING EFFECTIVENESS ---------- 25

A. TESTS ---------- 25

B. KOLMOGOROV-SMIRNOV TEST ---------- 25

C. SPEARMAN RANK CORRELATION COEFFICIENT ------- 31

D. SUMMARY ---------- 35

IV. MATRIX METHOD FOR THE EVALUATION OF TRAINING EFFECTIVENESS ---------- 36

A. INTRODUCTION ---------- 36

B. METHOD ---------- 37

C. ANALYSIS AND APPLICABILITY ---------- 41

V. A PARAMETRIC METHOD ---------- 44

VI. AN OPERATIONAL APPROACH TOWARD TRAINING EVALUATIONS ---------- 47

VII. A GENERAL METHODOLOGY ---------- 52

VIII. VALIDATION OF DATA COLLECTION METHODS ---------- 54

A. DATA COLLECTION ---------- 55

B. DATA ANALYSIS ---------- 56

C. TESTING FOR STATISTICAL SIGNIFICANCE -------- 57

D. ADDITIONAL INFORMATION ---------- 59

IX. SUMMARY AND CONCLUSIONS ---------- 61

APPENDIX A: CURRICULUM DESCRIPTION AND ASSOCIATED RANK REQUIREMENTS ---------- 63

APPENDIX B: HUMAN FACTORS IN TASK ANALYSIS ---------- 65

APPENDIX C: PERFORMANCE TASK QUESTIONNAIRE ---------- 74

APPENDIX D: CUMULATIVE STEP FUNCTION DATA IN KOLMOGOROV-SMIRNOV TESTS ---------- 77

APPENDIX E: TASK TRAINING AND JOB IMPORTANCE RANKINGS ---------- 80

LIST OF REFERENCES ---------- 83

INITIAL DISTRIBUTION LIST ---------- 84

FORM DD 1473 ---------- 85

LIST OF TABLES

Table

I Cumulative Distribution on Data Points in Kolmogorov-Smirnov Test with Basic Course Data ---------- 28

II Cumulative Step Functions in Kolmogorov-Smirnov Test with Basic Course Data ---------- 29

III Summary of Kolmogorov-Smirnov Test Results ------ 30

IV Summary of Spearman Rank Correlation Coefficient Computations ---------- 33

V Results of the Matrix Method of Evaluation ------ 41

VI Task Classification Criteria ---------- 42

VII Summary of Job Inventory Data Collected for Task Element Number Three ---------- 50

VIII Summary of Job Inventory Data Collected for Task Element Number Sixteen ---------- 50

IX Summary of the Sign Test Results on the Pairs of Data Collection Methods ---------- 58

LIST OF FIGURES

Figure

1 Hypothetical Representation of Job Proficiency Versus Training Emphasis ---------- 16

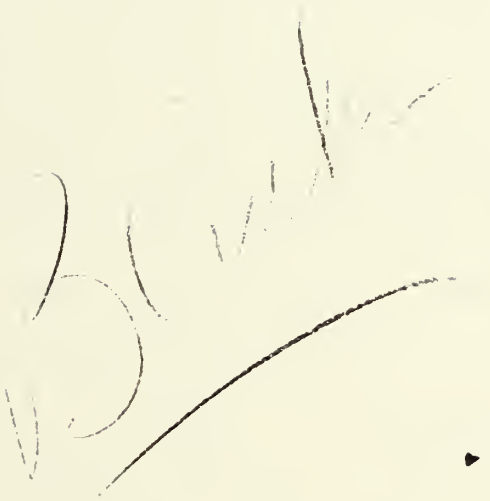
2 Hypothetical Representation of Job Proficiency Versus on the Job Training Emphasis ---------- 18

3 Matrix Scheme for Task Classification by Frequency of Performance and Training Level ---------- 38

4 Weights of Matrix Cells for Training Index Determination ---------- 38

5 Weights of Matrix Cells for Over Training Index Determination ---------- 40

6 Weights of Matrix Cells for Under Training Index Determination ---------- 40



I. INTRODUCTION

The increasing technology of military developments has been surpassing the military services' ability to supply competently trained manpower for meeting the knowledge and technical skill requirements for particular military jobs. While the adequacy of military training is dependent on the existence of effective procedures for developing and maintaining the appropriate instructional content, a means of judging the value or the worth of the training program has also become important. It is, however, quite misleading to associate any training model or training curriculum with a strict effective-ineffective judgement, since training programs may be considered adequate under certain circumstances and inadequate under others.

In an attempt to achieve a high positive correlation between job proficiency levels and the associated training costs, all of the military services are striving to establish instructional systems around a solid foundation of job data which portrays valid instructional objectives. Since job data refers to a statement of the work that an individual actually performs on the job, the ideal source for the collection of this data is the incumbent performing in his assigned billet. The technique of identifying and collecting these essential elements of the job is called task analysis. Basically, by utilizing a carefully structured

questionnaire containing a composite list of the probable tasks included in a job, as well as a rating scale for classifying tasks as to their degree of necessity, the extent to which certain elements of the job constitute an individual's performance can be precisely identified. Once a precise identification of the emphasis devoted to specific duties on the job is achieved, training objectives can be adjusted in order to bring training emphasis more on target. Task analysis, therefore, is useful in determining what training should be given and when it should be given in order to be timely.

Prior to the introduction of task analysis procedures, course contents were primarily constructed to include those areas of instruction which were thought to be required for adequate on the job performance. Additionally, no quantitative methods were available for evaluating training programs within the military services. With the advent of task analysis and the subsequent numerical taxonomy of job elements, sufficient data has become available for the utilization of statistical techniques in analyzing the effectiveness of military training.

This paper is based on the continued interest of the United States Marine Corps in the area of task analysis and job data utilization, as well as the recently initiated program of military occupational studies. In particular, the primary objective of the paper is to demonstrate the usefulness of task analysis in the evaluation of military

training programs. More specifically, a task analysis performed by the investigator is employed as a data base for illustrating some of the appropriate statistical techniques available for evaluating the effectiveness of three related Marine Corps training courses. Additionally, an approach toward determining the operational as well as the statistical significance of the effectiveness of these three courses is presented.

As a secondary objective, validation techniques and statistical comparisons are offered for three task analysis data collection methods applicable to the job inventory questionnaire. In this regard, the results of a small scale experiment designed to determine the most valid questionnaire associated data collection method are presented.

It is hoped that this research will allow quantitative as well as qualitative training curriculum analysis to become a more standard and definitive process within the Marine Corps.

A. SCOPE AND LIMITATIONS

One formal Marine Corps training program conducted at the Marine Corps Service Support Schools, Marine Corps Base, Camp Lejeune, North Carolina, is used to demonstrate the usefulness of task analysis in the evaluation of military training. The training program is subdivided into three separate courses designed to allow trainees to advance from one course to the next as they advance in rank. Basically,

the program is structured to include a basic, an intermediate, and an advanced level course, and is devoted to training Marines from the entry level rank through E-7 in the area of supply administration. A description of the subject curricula as well as the rank requirements for each is presented in Appendix A.

The data base to be used for illustrating the analysis techniques applicable to training evaluation was collected from a representative sample of Marines serving with MOS (3041) supply administrative man manual and MOS (3042) supply administrative man mechanized. Since the basic training course is designed to prepare Marines for filling manual supply billets, only Marines with MOS (3041) were used to provide data for the evaluation of this particular course. However, the intermediate and advanced level courses are devoted to preparing Marines for both manual and mechanized supply positions; therefore, Marines with MOS (3041) and MOS (3042) were employed to provide data for the evaluation of these two courses. The data has been aligned into three distinct categories representative of the Marine's most recent course completion. This arrangement provides data in an easily accessible manner for evaluating the individual courses as well as the overall training program.

The human variables contributing to successful job performance can be categorized into two specific areas of human factors oriented elements. One of these categorizations includes the actual units of work that form a

significant part of a job and are referred to as performance tasks or task elements. The other categorization includes the worker characteristics required for satisfactory job performance. These characteristics are not related to specific duties but include such factors as physical strength and mental requirements. A job inventory questionnaire was selected as the device for collecting job data and was designed to include these two categorizations of human factors oriented elements.

The performance task portion of the questionnaire consists of actual task elements which represent the essential part of the research. Data obtained from responses to these selected job description statements is used to demonstrate techniques for determining whether the skills needed for satisfactory performance in an administrative supply billet are consistent with the course contents of the associated formal training curriculum. The performance task portion of the questionnaire developed by the investigator is illustrated in Appendix B.

Although the primary emphasis of this paper is devoted to the analysis of actual performance task data, the worker characteristic segment of the questionnaire is representative of an attempt to quantify the human characteristics required for work in the supply field. This portion of the questionnaire was developed by Professor Gary K. Poock, a member of the Operations Analysis faculty at the Naval Postgraduate School and is displayed in Appendix C. Obviously, the data

obtained from the worker characteristic section of the questionnaire does not lend itself to training curriculum evaluation but was useful in providing an additional data base for the validation and comparison of different data collection methods. The information obtained from this particular portion of the questionnaire would be most appropriately applied to an analysis of MOS assignment procedures.

B. ASSUMPTIONS

The basis for the quantitative evaluation of training effectiveness presented in this paper rests on the assumption that the most important tasks performed by an individual on the job are the ones which occupy the major portion of his time. Accordingly, these elements of the job which occupy the most time are the ones in which students should receive the most intensive training. Basically, a student should be trained to perform those tasks which will be required of him in the field. Of course, the more specific the purpose of the training program in terms of performance requirements for particular military jobs, the more appropriately can the effectiveness of the program be determined by measuring the emphasis devoted to specific tasks by its graduates. The training objectives of many military schools are established to prepare graduates for specific job functions, while others are purposely constructed so that the dependence of job performance upon the training program is quite indirect.

More specifically, the approach to be taken in this paper implies a direct relationship between the training curriculum and the job. This approach represents the first disadvantage in the assumptions presented above in that this direct relationship is not characteristic of all training programs. Curriculums devoted to providing general background type knowledge, or those normally associated with higher level and more complex positions, often intentionally avoid a close relationship with specific elements of the job.

Another disadvantage associated with the assumptions presented above is related to the degree of difficulty inherent in each of the elements of the job. Skills which are not easily learned should, in most cases, receive more emphasis in the course content. In this regard, some elements of a job, although performed less frequently than others, require more intensive instruction to produce the desired proficiency level of the trainee. To eliminate this restriction, relationships between job proficiency levels and training emphasis could be established for the essential elements of the training program.

As an example of such a relationship consider the hypothetical curves shown in Figure 1. These curves represent the job proficiency level versus the training emphasis devoted to two maintenance oriented tasks. Since rotating truck tires is relatively simple and learned rapidly, maximum proficiency is achieved quite quickly. In contrast,

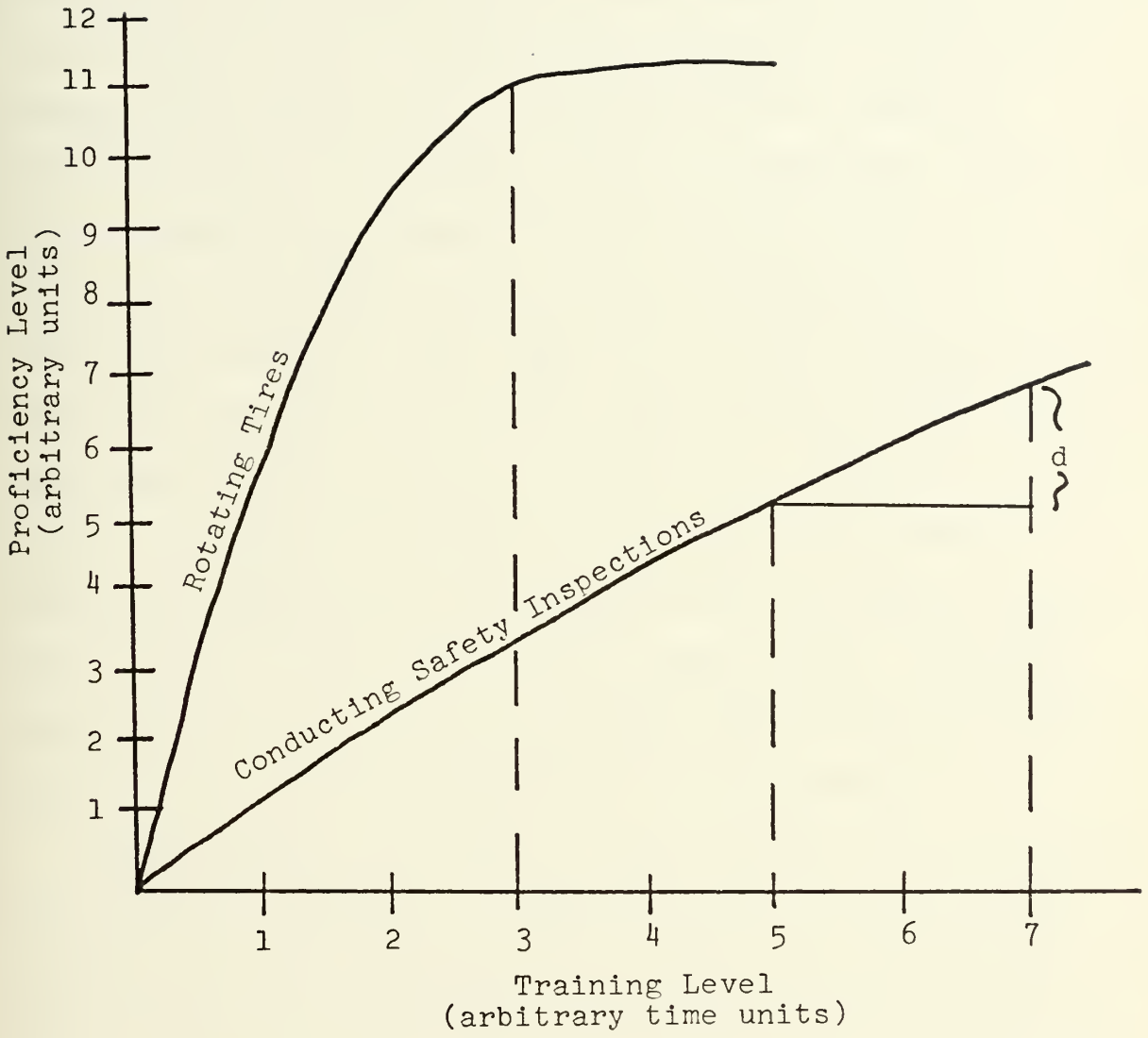


Figure 1. Hypothetical Representation of Job Proficiency Versus Training Emphasis

conducting vehicle safety inspections is a complex task which is learned more slowly. Accordingly, improvement in proficiency can be obtained from additional training time. As an example of this relationship, consider a training program which devotes five training units to rotating truck tires and five training units to conducting vehicle safety inspections. Suppose the program was developed on the basis of a survey demonstrating that approximately equal times are devoted to these tasks on the job. Inspection of the graph indicates that if two training units were eliminated from rotating tires, no reduction in proficiency would occur. These same two units of training could then be applied to increase the proficiency level of the more complex task by a significant amount equal to (d). Altering the curriculum in this way certainly represents an improvement in the overall effectiveness of the program.

Additionally, the aspect of task suitability for "on the job training" emphasis is representative of still another disadvantage with the assumptions employed in this paper. As an example of this aspect of training to the attainment of job proficiency levels, consider the hypothetical curves shown in Figure 2. The graph suggests that two independent tasks possibly occupying equivalent amounts of actual performance time are not equal in their applicability to training on the job. Task number 1 is relatively simple and indicates the achievement of a high proficiency level in a very short time period. In particular, this task is

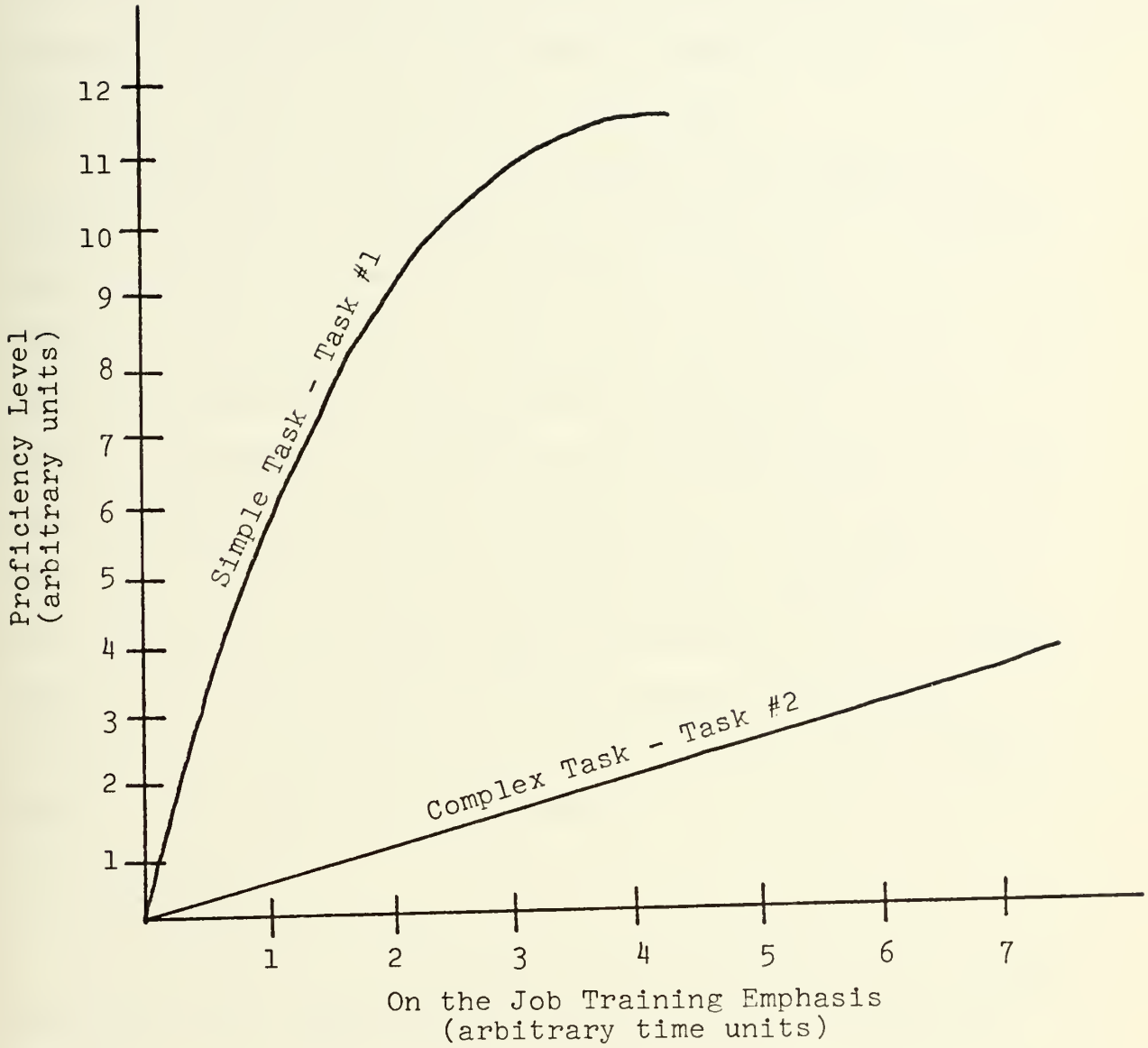


Figure 2. Hypothetical Representation of Job Proficiency Versus On the Job Training Emphasis

definitely suitable for on the job training and a likely candidate for elimination from the formal curriculum. Task number 2 is more complex and does not appear to be applicable to training techniques of this type.

Although the concepts just examined have substantial theoretical value and appear to be feasible on the surface, they represent several very difficult obstacles. The inclusion of these notions would require a variety of arbitrary decisions beyond the realm of this paper. It is, however, important to understand that deficiencies exist in the subsequent evaluation procedures if the previously mentioned complex issues are not included. Additionally, it is important to realize that accepting the basic assumptions will produce valid training program evaluations provided that qualitative considerations are applied subsequent to the quantitative analysis.

C. METHODOLOGY

The evaluation techniques to be demonstrated are based on statistical comparisons of the following two sets of data:

1. The relative importance of the tasks involved in the job.

2. The training emphasis devoted to the tasks involved in the job.

The first set of data was obtained by utilizing the task inventory questionnaire in conjunction with a personal interview and actual observation of the incumbents' performance

on the job. Basically, individuals were required to report the extent to which the tasks listed on the questionnaire occupied their time. In keeping with the basic assumption of this paper, those tasks which occupied the most time were considered to be the most important. To limit the selection of responses available for reporting the extent to which specific tasks occupied their time, a four point selection scale was used. Incumbents considered each task with the investigator and were required to explain their selected response from the four point scale. It was anticipated that the required explanation would insure a complete understanding of each question, while at the same time limit the workers from over responding. The questionnaire employed and the associated scoring procedure is illustrated in Appendix C. Scores were converted to percentages to provide more meaningful data in nature of the comparisons to be made. A computer program was developed to consolidate and display the data in an orderly manner. The task analysis questionnaire was administered to 95 enlisted Marines ranking from E-2 through E-8. Once collected, the data was classified in accordance with the Marines' most recent course completion. As previously mentioned this classification provided data in an easily accessible manner for evaluating the three courses as well as the overall program. The results of this classification provided 29 subjects for the basic course evaluation, 50 subjects for

the intermediate level course analysis, and 16 subjects for the evaluation of the advanced course.

The second set of data was obtained from theoretical proficiency levels specified by the directors of the training program.

Comparisons of these two sets of data will determine if the training program is devoted to placing emphasis on the appropriate instructional areas. Statistical techniques are used to illustrate some of the available procedures for evaluating the three courses on an individual basis. The results of these individual course evaluations are then extended in order to demonstrate methods for determining the effectiveness of the overall program.

D. TASK ANALYSIS DATA COLLECTION METHODS EMPLOYED

Although a variety of task analysis data collection methods exist, those employed in this study are representative of the procedures applicable to the job inventory questionnaire. The questionnaire permits extensive sampling of military occupational specialties in a relatively short period of time. Additionally, the questionnaire oriented procedures are economically administered and allow large masses of data to be reduced for computer storage, display and analysis. For these reasons it is estimated that the data collection methods employed, although far from exhaustive, are representative of those feasible for Marine Corps efforts in this area. The following are the task analysis job inventory procedures utilized in this paper:

1. Personal interview and actual observation of an individual's on the job performance.

2. Distribution of the task analysis inventory questionnaire in a manner not permitting observation or interview of the job incumbent. This procedure requires individual completion of the questionnaire with no explanation of the task statements.

3. Utilization of the task analysis questionnaire by officers and senior staff non-commissioned officers serving in a supervisory capacity. For this method job data was obtained by supervisors reporting the job characteristics required of their subordinates.

Only the data collected using the questionnaire in conjunction with a personal interview and observation of the job incumbent was used in the evaluation procedures demonstrated in this paper. The other data gathering procedures were useful in determining the most valid method for acquiring job inventory data.

II. STATISTICAL TESTING IN GENERAL

The purpose of a training course evaluation is to provide information regarding the effectiveness of the curriculum content. Of particular interest here is to examine alternative approaches toward determining whether or not training is being conducted in the appropriate instructional areas and if students are of the proper rank in order for the training to be timely. As an example of the concept of timely training, consider a motor transport mechanic who at some time in his career may be required to employ management techniques in the operation of a large repair facility. Since these management techniques will probably only be required of senior enlisted men, a timely training program would not provide instruction in this area for the entry level or even the intermediate level ranks. What seems to be needed here is a good method for determining the effectiveness of course contents while at the same time providing a foundation capable of incorporating the ideas mentioned above into an evaluation of the overall program. Possible approaches toward achieving this method include the use of nonparametric as well as parametric statistical techniques.

A parametric statistical test is a test in which specific assumptions about the parameters of the sampled population are made, such as $\mu_1 = \mu_2$ or $\sigma_1 = \sigma_2$, whereas a nonparametric test makes no assumptions about the value

of the parameters in the sampled population. It is important to distinguish the difference in the assumptions which must be made when testing parametrically vice nonparametrically. The basic assumptions of the parametric test include independence of observations, underlying normal distribution of the sampled populations, and homoscedacity of the population variances. The assumptions associated with nonparametric tests include only the independence of observations, and that the sampled populations be continuous. When judged by the criterion of power efficiency, nonparametric tests are often superior to their parametric counter parts when all of the assumptions required of the parametric tests can not be met [1].

In keeping with the primary objective of this paper, which is demonstrating the usefulness of task analysis in the evaluation of military training, an attempt is not made to determine which test or even which type of test is the best for such an analysis. Rather than make such a determination, a variety of applicable evaluation procedures are presented. In addition to presenting applicable statistical evaluation procedures, the concept of operational effectiveness and a general methodology for standardizing military training evaluations are examined.

III. NONPARAMETRIC EVALUATIONS OF TRAINING EFFECTIVENESS

A. TESTS

The applicable tests selected to demonstrate nonparametric statistical procedures for evaluating the effectiveness of the Marine Corps training courses are the Kolmogorov-Smirnov Test and the Spearman Rank Correlation Coefficient. A matrix method of evaluation, also a form of nonparametric testing, will be examined in a separate chapter.

B. KOLMOGOROV-SMIRNOV TEST

Perhaps the most huerestic of the statistical tests is the Kolmogorov-Smirnov One Sample Test, often referred to as the Smirnov Maximum Deviation Test. The test statistic is the maximum deviation between two empirical distribution functions [4]. Since training effectiveness in this paper is to be determined by evaluating the extent to which course contents are job oriented, the actual proportions of job time devoted to the different task elements are used as the standard. Therefore, the distribution of job time over the task elements is considered the theoretical fixed cumulative distribution. Basically the data in each category, i.e., basic, intermediate, and advanced, was collected using the four point rating scale displayed in Appendix C and was then individually averaged over every task statement on the questionnaire. These averages were then converted to percentages in order to provide the average proportion of time

devoted to each task element by the most recent graduates of each of the three Marine Corps training courses. More specifically, a fixed theoretical cumulative distribution of job time was derived for each of the three categories of actual job data.

The other empirical distribution was determined from the directors of the training program and represents the proportion of time devoted to the task elements within each of the three courses. So at this stage of the procedure, a theoretical fixed cumulative distribution of actual job time and a cumulative distribution of training emphasis were available for each of the training courses.

To compute the test statistic, list the values of the cumulative distribution functions of training emphasis and job time in order of the various task elements. The test statistic, D, is $\max|d_i|$. Where:

$$d_n = \sum_{i=1}^{n} J_i - \sum_{i=1}^{n} T_i, \quad i = 1,2,\ldots n,\ldots 32,$$

and J_i = average proportion of actual job time devoted to task element i.

T_i = proportion of training time devoted to task element i.

Testing is conducted under the null hypothesis, H_o:, that the two distributions are the same and the alternative hypothesis, H_1:, that the two distributions are not the same.

It is important to note that the Kolmogorov-Smirnov Test used in the manner explained above does not meet all of the

prerequisites required of nonparametric statistical tests. The data is cast into intervals defined by each of the tasks listed on the questionnaire, and therefore, cannot meet the continuity assumption required for this nonparametric test. Additionally, the results of the Kolmogorov-Smirnov Test used in this way produce a somewhat pessimistic appraisal of the effectiveness of the training courses, since the value of D will be at least as great as the maximum difference between the job time versus training emphasis on any one task element. The test, however, has intuitive appeal in that it considers proportional differences on distinct task elements as well as runs of task statements which display discord between the respective proportions.

As an illustration of the manner in which this test was employed, consider the data associated with the Basic Supply Course. An explanation of the nature of this course is presented in Appendix A. The data points of the two cumulative distributions as well as the d_i's are shown in Table I. Similar tables for the remaining courses are presented in Appendix D. Table II illustrates a general plot of the two cumulative step functions associated with the Basic Supply Course.

TABLE I

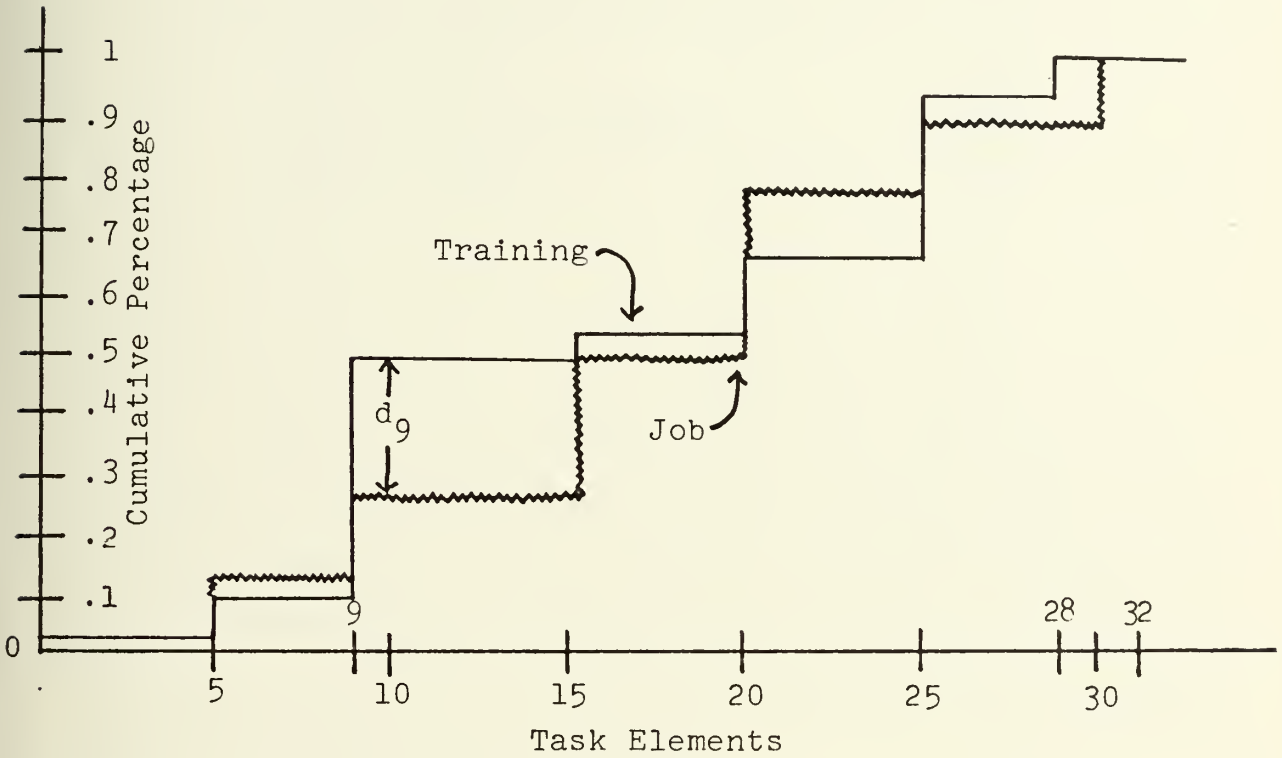
CUMULATIVE DISTRIBUTION DATA POINTS IN KOLMOGOROV-SMIRNOV TEST WITH BASIC COURSE DATA

Task No.	Training (CDF)	Job (CDF)	d_i
1	.02	.0018	.0182
2	.02	.0036	.0164
3	.02	.0250	.0050
4	.02	.0250	.0050
5	.09	.1015	.0115
6	.10	.1086	.0086
7	.28	.1816	.0984
8	.40	.2688	.1312
9	.51	.2955	.2145
10	.56	.3346	.2254*
11	.56	.3666	.1934
12	.56	.3755	.1845
13	.56	.3986	.1614
14	.56	.4591	.1009
15	.57	.5036	.0664
16	.60	.5890	.0110
17	.60	.6566	.0566
18	.60	.6993	.0993
19	.61	.7456	.1356
20	.67	.7634	.0934
21	.68	.7794	.0994
22	.73	.8257	.0957
23	.92	.8542	.0658
24	.96	.8791	.0809
25	.96	.8791	.0809
26	.96	.8846	.0754
27	.99	.9273	.0627
28	1.00	.9842	.0158
29	1.00	1.0000	0
30	1.00	1.0000	0
31	1.00	1.0000	0
32	1.00	1.0000	0

*D = .2254

TABLE II

CUMULATIVE DISTRIBUTION STEP FUNCTIONS IN KOLMOGOROV-SMIRNOV TEST WITH BASIC COURSE DATA



D occurs at i = 9 where

$$d_9 = .56 - .3346 = .2254.$$

When compared with the critical value of .1613 it is possible to reject the null hypothesis at a significance level of .01. The values of D for the courses under examination are shown in Table III. The null hypothesis when applied to the remaining courses may also be rejected at the .01 significance level.

A rejection of the null hypothesis at the .01 significance level for all three courses indicates that training

TABLE III

SUMMARY OF THE KOLMOGOROV-SMIRNOV TEST RESULTS

	D
Basic Course	.2254
Administrative Course	.1824
Senior Course	.2736

emphasis is not devoted to the appropriate instructional areas. However, more important than the statistical significance determined by the Kolmogorov-Smirnov Test is the manner in which the test presents the data. The examination of the two cumulative distributions from a more operational standpoint, may indicate feasible adjustments for bringing training emphasis more on target.

As an example of some of the comparisons that can be made, consider the illustration of the two cumulative distribution functions presented in Table II. Tasks nine through fifteen represent a trend toward overtraining. The over emphasis devoted to training on these tasks may be more appropriately applied to tasks twenty through twenty five, where an undertraining trend is indicated. Similar comparisons can be made among the three courses in order to locate possible adjustments for improving the training program on an overall basis. More elaborate comparisons could be made for evaluations containing a much greater number of performance tasks, which could be classified and sequenced into

distinct instructional areas. Nevertheless, the Kolmogorov-Smirnov Test, as illustrated, is useful in providing an indication of the effectiveness of the training courses.

C. SPEARMAN RANK CORRELATION COEFFICIENT

Of all the statistical tests based on ranks, the Spearman Rank Correlation Coefficient is perhaps the oldest and the most widely known. The test statistic r is a measure of the degree of association that exists between two groups of data [5].

To compute r it is necessary to establish two separate categories of data based on the assignment of ranks to the tasks listed on the questionnaire. The first ranking is based on the relative importance of the tasks to satisfactory performance on the job while the second is achieved from the emphasis devoted to the tasks in the training program. The task determined the most important to the job is assigned a rank of 1 for the job data category while the task receiving the most intensive training emphasis is assigned a rank of 1 for the training data category. The remaining tasks are ranked accordingly.

The test statistic

$$r = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N^3 - N}$$

where:

d_i = the difference between the two ranks assigned to task element i.

N = the number of task elements.

The value of r ranges between 0 and 1. The highest possible degree of association is 1, and can be achieved if, and only if, each task element receives the identical ranking within the two categories of data.

If the proportion of ties within the rankings of either category is large, it becomes appropriate to adjust the value of r by a correction factor. The correction factor,

$$T_j = \sum_{i=1}^{n} \frac{t_i^3 - t_i}{12}, \quad j = X,Y$$

where:

t_i = the number of observations tied at rank i.

n = the total number of ranks at which tied observations exist within a given category of data.

The variable X represents the training data, while the variable Y represents the job data. With the correction factor incorporated one can use the formula:

$$\bar{r} = \frac{X^2 + Y^2 - \sum_{i=1}^{N} d_i^2}{2\sqrt{X^2 \cdot Y^2}}$$

where:

$$X^2 = \frac{N^3 - N}{12} - T_X$$

$$Y^2 = \frac{N^3 - N}{12} - T_Y.$$

The rankings required for the computation of r for the curricula under observation are listed in Appendix E.

Additionally, it is possible to test the significance of the value of r or $\bar{r}$. This can be done by testing the null hypothesis, Ho: that the two variables under observation are not associated against the alternative hypothesis, H_1: that there is a measure of association between the two variables. The number of task elements being greater than ten determines the use of the significance test for large samples [5]. The significance of an obtained r under the null hypothesis may be tested by:

$$t = r\sqrt{\frac{N-2}{1-r^2}}$$

The test statistic, t, has the student's t distribution with degrees of freedom = N - 2, for large values of N.

Table IV illustrates a summary of the results obtained by using the Spearman Rank Correlation Coefficient.

TABLE IV

SUMMARY OF SPEARMAN RANK CORRELATION COEFFICIENT COMPUTATIONAL RESULTS OF JOB TIME VERSUS TRAINING EMPHASIS

	r	$\bar{r}$	t
Basic Course	.557	.531	5.446
Administrative Course	.282	.279	1.60
Senior Course	.366	.357	2.140

Since the values of r are so closely related to $\bar{r}$, the significance test was only conducted on the uncorrected measures

of association. The results indicate that the null hypothesis can be rejected in the following instances:

1. Basic Course at a significance level of .001;
2. Administrative Course at a significance level of .20;
3. Senior Course at a significance level of .05.

A rejection of the null hypothesis indicates that a measure of association does exist between training emphasis and job importance for the tasks listed in the questionnaire. Therefore, the values of r computed for the three training courses are statistically significant at the levels shown above.

In addition, qualitative analysis is required here in order to determine if the values of r are close enough to 1 in order for the training program to receive an effective rating. The establishment of a standard more easily obtainable than a correlation coefficient of 1 may prove appropriate in order for this test to produce a clear evaluation of the effectiveness of training. Although such an extension of this test would certainly provide more meaningful results, it is beyond the realm of this paper to attempt even an approximation of this standard value.

The Kendall Rank Correlation Coefficient is another measure of association type test that is appropriate here but will not be demonstrated in this paper.

D. SUMMARY

This concludes a presentation of possible nonparametric statistical tests which could be used in evaluating the effectiveness of military training. The results of the two tests conducted in this chapter appear to be conflicting. The Kolmogorov-Smirnov Test suggested that training emphasis was not being devoted to the proper instructional areas while the Spearman Rank Correlation Coefficient indicated a degree of association between training emphasis and job importance levels. While it is difficult to explain the cause of these differences, it is certainly safe to say that the assumptions of the Spearman Rank Correlation Coefficient are most closely adhered to while the violation of the continuity requirement in the Kolmogorov-Smirnov Test may tend to discount the statistical validity of its results.

IV. A MATRIX METHOD FOR THE EVALUATION OF TRAINING CURRICULUM EFFECTIVENESS

A. INTRODUCTION

The matrix method of training curriculum analysis presented here has been adopted from a study conducted by Arthur I. Siegel, Douglas G. Schultz, and Phillip Federman in 1961 for the Office of Naval Research [7].

This correlation oriented procedure enables the analyst to determine a training index, an undertraining index, and an overtraining index for each curriculum under evaluation. The numerical value of each index will be a number greater than or equal to zero and less than or equal to one. The training index is representative of the overall effectiveness of the program while the overtraining and undertraining indexes represent the direction of problem areas if the curriculum is classified as ineffective. A training index of one represents the highest possible positive correlation between adequate job performance and the associated task proficiency levels established by the training program. Basically, an effective curriculum content should produce a training index close to one while the undertraining and overtraining indexes should be near zero.

Although the matrix method requires some value judgement as to the placement of tasks in the matrix cells, its basic advantage is the relative ease with which calculations can be accomplished.

B. METHOD

The procedure is based on the categorization of tasks into four groups labeled "very high," "high," "moderate," and "low." Separate classifications are conducted for the training level associated with each task as well as the relative importance of each task to performance on the job. Subsequent to these classifications each task is then sorted into one of the sixteen cells in the matrix shown in Figure 3. Cell A, for example, contains tasks which are very highly important and for which trainees have been trained to a very high proficiency level. Cell H contains tasks which are greatly important but received little training emphasis. By assigning weights to the cells of the matrix, the three indexes can be calculated as follows:

1. Training Index (TI): If all tasks were instructed to a level equivalent to that required for job performance, one would anticipate all of the tasks to locate in cells A,F,K, and P. These cells are assigned a weight of three. Cells B,G,L,E,J, and O are assigned the weight of two, cells C,H,I, and N are assigned the weight of one, while cells D and M are assigned the weight of zero. Figure 4 illustrates the assignment of weights to the cells of the matrix.

The urriculum Training Index is determined by the sum of all the weights associated with the tasks of a particular curriculum divided by three times the number of tasks that are being evaluated. The mathematical expression is as follows:

Frequency of Performance

Training Level	Very High	High	Moderate	Low
Very High	A	B	C	D
High	E	F	G	H
Moderate	I	J	K	L
Low	M	N	O	P

Figure 3. Matrix Scheme for Task Classification by Frequency of Performance and Training Level

Frequency of Performance

Training Level	Very High	High	Moderate	Low
Very High	3	2	1	0
High	2	3	2	1
Moderate	1	2	3	2
Low	0	1	2	3

Figure 4. Weights of Matrix Cells for Training Index Determination

$$TI = \frac{\sum_{i=1}^{N} W_i}{3N}$$

where:

W_i = the weight assigned to the cell in which task i has been classified.

N = the number of tasks being classified.

2. Overtraining Index (OI): The realization of a relatively low training index indicates the appropriateness of measuring the direction of ineffectiveness in the training program. Basically, this means a determination of the extent to which individuals are being overtrained or undertrained.

A measure of the tendency toward overtraining can be determined by the assignment of weights to the matrix cells as shown in Figure 5. Only the cells that indicate overtraining have been assigned a weight. Computationally, the overtraining is as follows:

$$OI = \frac{\sum_{i=1}^{N} W_i}{3N}$$

Ideally, an effective training curriculum will produce an overtraining index of zero.

3. Undertraining Index (UI): This index represents a measure of the tendency toward undertraining and is determined in accordance with the principles shown above. Matrix cells are weighted as illustrated in Figure 6. Only the

Training Level	Frequency of Performance: Very High	High	Moderate	Low
Very High	0	1	2	3
High	0	0	1	2
Moderate	0	0	0	1
Low	0	0	0	0

Figure 5. Weights of Matrix Cells for Overtraining Index Determination

Training Level	Frequency of Performance: Very High	High	Moderate	Low
Very High	0	0	0	0
High	1	0	0	0
Moderate	2	1	0	0
Low	3	2	1	0

Figure 6. Weights of Matrix Cells for Undertraining Index Determination

cells indicating undertraining have been assigned a weight. Computationally, the Undertraining Index is as follows:

$$UI = \frac{\sum_{i=1}^{N} W_i}{3N}$$

The three indexes for each training course, taken together, provide a useful summary of training effectiveness in each instructional area. While the indexes themselves provide a valuable indication of the effectiveness of training, the task classification procedure is more specific in pointing out where increases and decreases in training emphasis are required.

C. ANALYSIS AND APPLICABILITY

As an application of the matrix method of evaluation, the three indexes were calculated for each of the courses under study. The results are shown in Table V. The training level classification for each task was provided by

TABLE V

RESULTS OF THE MATRIX METHOD OF EVALUATION

Course	Training Index	Overtraining Index	Undertraining Index
Basic	.802	.073	.125
Administrative	.813	.094	.094
Senior	.823	.125	.052

the directors of the training program, while the frequency of performance level was determined in accordance with the criteria displayed in Table VI.

TABLE VI

TASK FREQUENCY OF PERFORMANCE CLASSIFICATION CRITERIA

Classification	Percentage of Time Devoted to Job
Very high	15 and up
High	6 - 14.99
Moderate	2 - 5.99
Low	0 - 1.99

For the training courses being evaluated, the three indexes each view the data in a different manner. When considered collectively, they provide a thorough indication of the extent to which the training program is preparing its graduates for performance on the job. No one index can be used to determine absolutely if one course is more effective than another, nor is any one index necessarily the one by which a particular program should be evaluated. Basically, all three indexes must be considered together and in relation to one another.

Without complete knowledge of the training objectives and the administrative constraints associated with the training program, it is almost impossible to assign a standard which the indexes should have met in order for the

courses to be considered effective. However, from a more academic standpoint, the deviation of the training index from 1 and the undertraining and overtraining indexes from 0 seem to be rather small for the three courses being examined. Therefore, a tendency toward an effective rating is suggested, although an effective-ineffective judgement at this stage of the analysis is somewhat impractical.

A thorough examination of each task located off the main diagonal in the classification matrix would have to be conducted in order to determine if adjustments in training emphasis would be feasible. Nevertheless, the training indexes as presented increase the understanding of the effects of the training program and provide a quantitative basis for altering training emphasis.

V. A PARAMETRIC METHOD

A possible parametric procedure for determining the quality of the training program is the student's "t" test. The usefulness of the test is limited to an effectiveness evaluation on each individual task element. Basically, it allows a statistical test to be conducted on the extent to which each task element of a training curriculum is job oriented. From the calculation of test statistics for each task statement and the subsequent determination of statistical significance, one could make an accurate analysis of the emphasis of the training content. Because of the large number of calculations involved, it is beyond the scope of this paper to employ the parametric procedure mentioned above. This chapter will be devoted to an explanation of the procedure and the assumptions that are required for using the "t" test in a meaningful way.

The problem becomes one of testing the hypothesis that the mean of a population equals a specified value against the alternative that the population mean is not equal to the specified value. The hypothesis is expressed as follows:

$$H_o: \quad \mu_i = \mu_{oi}$$

$$H_1: \quad \mu_i \neq \mu_{oi}$$

where:

μ_i = mean importance level of task i

μ_{oi} = established proficiency level associated with task i in the training curriculum.

The test is based on the reasonableness of the assumption that the population of importance levels are normally distributed [2]. If this assumption is feasible within a reasonable degree of approximation, then the test statistic is as follows:

$$t_{obs} = \frac{\bar{X}_i - \mu_{oi}}{s/\sqrt{n}}$$

where:

$\bar{X}_i$ = the mean of the sampled population

s = the variance of the sampled population

n = the number of job incumbents from which data was collected.

The critical value is extracted from any standard "t" table and is based on n-1 degrees of freedom as well as the desired level of significance. If t observed is greater than t critical, the null hypothesis is rejected and it may be stated that training is not job oriented for the task element tested.

This same procedure is then repeated for each element of the program so that either an acceptance or rejection of the null hypothesis is accomplished for every task statement.

The t test used in this manner has the advantage of evaluating the significance of every element of the program

and provides accurate statistical information for altering the emphasis of course contents. Since a large number of task elements imply extensive calculations, computer techndiques are suggested for computational efficiency.

VI. AN OPERATIONAL APPROACH TOWARD TRAINING EVALUATIONS

Previous chapters of this paper discussed nonparametric and parametric procedures for evaluating the effectiveness of military training. Alternative approaches were illustrated providing statistical techniques for analyzing the extent to which training is job oriented. These procedures provide insight into the emphasis of course contents based upon proven statistical methods. However, it still remains impractical to assign an effective-ineffective rating to any training curriculum, and therefore, evaluations based on the operational significance of the data are also required.

As a possible approach toward an appraisal of the training program in a more operational manner, it appears appropriate to consider the classification procedure demonstrated in the chapter on the matrix method of evaluation. The statistical procedures demonstrated in Chapter III may provide equally acceptable starting points; however, only the matrix method will be pursued here. Tasks located off the main diagonal of the classification matrix are representative of candidates for a more intensive examination of proper training emphasis. Those tasks assigned to cells D,C,H,I,N, and M of the matrix should, therefore, be of particular concern to the directors of the training program.

Subsequent to determining candidate tasks for a more detailed analysis on an individual basis, it seems

appropriate to examine the job data for trends or patterns based on the nature of the selected tasks. Comparisons of these trends or patterns associated with the importance of the tasks to performance on the job against those related to training emphasis provide information as to the timeliness of the training program. More specifically, it is not only important to determine what training should be given, but when it should be given in order to be timely. Discrepancies at this stage of the analysis indicate an untimely training program where individuals are either being trained prematurely or subsequent to the time that they are required to perform the task.

As an example of the timeliness concept, consider a complex task not normally required of lower ranking individuals but continuously becoming more important as the individual increases in rank. A timely training program would not provide detailed instruction on this task for the entry level ranks but would emphasize the task within the more advanced courses.

Alterations in training emphasis indicated by an untimely training program would include the elimination of instructional areas from one course while adding them to another. In keeping with the objective of military training, these additions and deletions should be accomplished in accordance with the importance of the tasks to be performed on the job. Other tasks receiving timely training may still require modifications in emphasis in order to

bring the program more on target. Adjustments in course contents, when the timeliness of training is not an issue, require an examination of the courses on an individual basis since a task receiving proper emphasis in one course may be overemphasized or underemphasized in another. More specifically, the evaluation procedure suggested here is associated with determining when training should be accomplished to be timely as well as determining the proper emphasis that should be devoted to specific instructional areas within individual courses.

As an example of the concepts discussed above, consider task Numbers Three and Sixteen. The tasks are listed on the job inventory questionnaire displayed in Appendix C. Task Number Three is "the assignment of work to personnel on a daily basis." The data also indicates that the frequency of performance of this task is continuously increasing as individuals advance in rank. Because task Number Sixteen is more menial in nature, its associated importance level decreases in accordance with the rank structure.

As previously mentioned, the job inventory data was separated into three distinct categories representative of the Marine's most recent course completion. Table VII illustrates the range of ranks associated with each category as well as the average percentage of time devoted to task Number Three on the job. Table VIII shows the average percentage of time devoted to task Number Sixteen within the same three categories of data.

TABLE VII

SUMMARY OF JOB INVENTORY DATA COLLECTED FOR TASK NUMBER THREE

Data Category	Rank	% Job Time
Basic	E2-E3	2.14
Intermediate	E4-E6	4.47
Senior	E7-E8	7.23

An observation of the curriculum contents reveals the same continuous tendencies as illustrated in Table VII and Table VIII. These results are indicative of a timely training program in which individuals are instructed at the proper rank for adequate job performance.

TABLE VIII

SUMMARY OF JOB INVENTORY DATA COLLECTED FOR TASK NUMBER SIXTEEN

Data Category	Rank	% Job Time
Basic	E2-E3	8.54
Intermediate	E4-E6	6.62
Senior	E7-E8	3.31

In contrast, task Number Eleven as listed on the questionnaire displayed in Appendix C, involves the maintenance of financial records. The data shows that this task is

performed more regularly by lower ranking individuals while training emphasis is more concentrated toward the senior enlisted man. Task Number Eleven is therefore representative of an untimely aspect of the training program and should receive a more intensive examination.

Once tasks have been evaluated for timeliness as demonstrated above, it still remains to examine the emphasis devoted to them within each individual training course. This portion of the analysis involves the modification of course contents in accordance with the importance of tasks associated to performance on the job and could be accomplished by adjusting training emphasis enough to relocate these tasks on the main diagonal of the classification matrix.

However, it still remains to consider the administrative type constraints associated with formal military training prior to the adjustment or reorganization of course contents. Since these constraints are obviously varied and unique to specific training programs, they will not be included here but are considered an appropriate area for future research. Although a specific judgment of the effectiveness of the program remains impractical, a more operational approach toward the employment of task analysis in the evaluation of military training has been introduced.

VII. A GENERAL METHODOLOGY

This chapter introduces a general systematic methodology for utilizing task analysis in the evaluation of military training. The major features of the proposed methodology have emerged from modern training research and include statistical techniques, operational evaluations, and comparisons of data collection methods. The procedure consists of ten steps to training curricula evaluation:

1. Conduct a systems analysis of the program to be evaluated, thereby determining the rank and MOS requirements of the individuals to be employed as subjects.

2. Determine the size and location of a representative sample of job incumbents from which job specification data can be obtained and consider sample stratification techniques if infeasible to include the entire population.

3. Develop the task analysis inventory to be administered to job incumbents. The foundation for this development comes from the systems analysis of the training program as well as established job requirements.

4. Select the data collection method to be used and conduct the task analysis.

5. Select the appropriate statistical procedure.

6. Consolidate the job data so that a comparison with the training proficiency levels can be conducted in accordance with the statistical technique selected.

7. Conduct the statistical test and determine if statistical significance exists.

8. Examine available data from an operational standpoint to include a consideration for the timeliness of training.

9. Adjust the training program where feasible.

10. Monitor graduates on a pericdic basis in order to maintain task analysis data in a current status.

The methodology suggested here is certainly general in nature and presents no revolutionary ideas in the search for an efficient way to evaluate training. It does, however, provide some procedural guidance for using task analysis in the evaluation and revision of military training curricula.

VIII. VALIDATION OF DATA COLLECTION METHODS

As mentioned in Chapter I, a secondary objective of this paper is to attempt, on a relatively small scale, statistical comparisons of some of the data collection methods associated with the job inventory questionnaire. This particular aspect of task analysis will be used to determine if the data obtained from any one collection method is significantly different from that obtained by one of the other procedures. Chapter I also lists the methods of data collection most feasible for Marine Corps task analysis efforts. Three of these four methods were subjected to a test to determine the degree of association existing between the three distinct combinations of data gathering procedures. The task analysis data collection methods selected were the most feasible for this type of comparison and are as follows:

1. Method Number 1 - Personal interview and actual observation of an individual's on the job performance. The investigator completed the questionnaire based upon verbal responses of the job incumbents.

2. Method Number 2 - Distribution of the task analysis questionnaire in a manner not permitting observation or interview of the job incumbent. The workers completed the questionnaire without assistance from a trained observer.

3. Method Number 3 - Distribution of the questionnaire to supervisors reporting on the job specifications required

of their subordinates. Supervisors completed the questionnaire without direct assistance from a trained observer.

A. DATA COLLECTION

The data for this segment of the paper was gathered over a period of about six weeks. Fifteen Marines from the Marine Corps Supply Center, Albany, Georgia, were employed as subjects. The Marines held a rank of either E-3 or E-4.

Initially, staff non-commissioned officers and officers completed the inventory questionnaire on the job characteristics that they felt were required of their men. The questionnaire was at the same time distributed to the fifteen job incumbents in accordance with data collection method Number Two. The subjects received a brief explanation of the instructions and were then given a period of about twenty-four hours to complete the questionnaire. No explanation of the individual task statements was permitted although precautions were not taken against subjects discussing some of the statements or procedures prior to marking the questionnaire.

Finally, after a period of about six weeks, the subjects were interviewed on the job. For this portion of the experiment the subject considered each task statement on the questionnaire with the investigator and actually was required to explain why he felt that particular tasks were required in the performance of his job. The investigator then marked the questionnaire in accordance with the response of the job incumbent.

The same questionnaire was employed throughout the test and is shown in Appendix C. In order to limit the selection of responses in regard to the extent to which specific tasks are required for performance on the job, a rating scale containing only four points was used. An explanation of the four point selection scale is also presented in Appendix C.

Additionally, a questionnaire useful in evaluating the extent to which human factor characteristics or worker characteristics are required for successful performance on the job was administered to these same subjects at the same time as the performance task questionnaire. This particular questionnaire was explained in Chapter I and was developed by Professor Gary K. Poock, a member of the Operations Analysis faculty at the Naval Postgraduate School. This questionnaire is displayed in Appendix B.

B. DATA ANALYSIS

Observation of the completed questionnaires revealed the development of an apparently significant trend in regard to the three data collection methods. The distribution method, in which individual workers were required to complete the questionnaire on their own, without detailed explanations of the instructions, yielded higher total point counts in all cases when compared with the investigator conducted interview and a higher total point count in ten of the fifteen cases when compared with the supervisors reports.

Additionally, the results of the questionnaire when used by supervisors to report on their subordinates (data collection method Number Three) demonstrated a higher total score for fourteen of the fifteen subjects when compared with the job interview method administered by the investigator. Total point count was obtained by adding the responses of each subject on both of the questionnaires that were used in each phase of the test.

C. TESTING FOR STATISTICAL SIGNIFICANCE

The nonparametric Sign Test was used to test for statistical significance between the pairs of data gathering procedures. Because of the large sample size of task statements the normal approximation to the binomial distribution was required for computation of the test statistic. For a detailed discussion of the rationale and method of this test see [6]. Three separate test statistics were calculated so that the following statistical comparisons could be made:

1. Method 1 versus Method 2
2. Method 1 versus Method 3
3. Method 2 versus Method 3

The following null hypothesis and alternative hypothesis were used throughout the testing:

Ho: there is no significant difference between the data gathering methods.

H_1: there is a significant difference between the methods.

The results of the tests are indicated in Table IX.

TABLE IX

SUMMARY OF THE SIGN TEST RESULTS ON THE PAIRS OF DATA COLLECTION METHODS

Test	Critical Value	P Value
1 vs 2	3.38	.0010
1 vs 3	3.02	.0026
2 vs 3	1.53	.1260

Statistically, the results of this test are determined by comparing the P value with the desired significance level (α). The null hypothesis can be rejected if P is less than or equal to the selected significance level. As an example, if $\alpha = .005$ is chosen, the results indicate a significant difference between data collection method Number One and the other two procedures. No statistical significance is demonstrated between methods Two and Three. Statistically, the results indicate that job data obtained by distributing questionnaires to job incumbents is not different from that collected by supervisors using the questionnaire to report on the job specifications required of their subordinates. Additionally, the results of the Sign Test indicate that the data obtained by personal interview and actual observation of the job incumbents is different from that obtained using the other two methods. The Sign Test, however, does not indicate which data collection method produces the most valid information

D. ADDITIONAL INFORMATION

In addition to the statistical results of the Sign Test, it is of interest to consider some specific items on the worker characteristic questionnaire in an attempt to determine which of the tested data collection methods produce the most valid information. Since this questionnaire was designed for use on all MOS's, some questions are completely irrelevant to the type jobs considered for this evaluation. Of particular interest is Question Twenty-nine which querries the subject as to his need for an accurate sense of taste, and Question Fifty-five which asks the worker if the ability to cause subordinates to willingly produce desired results is required in the performance of his job. These particular worker characteristics are representative of those definitely not required of any of the individuals involved in the testing. None of the job incumbents was in any way associated with tasting food in the performance of his job. Additionally, none of the subjects had the responsibility of subordinates nor were they required to possess the capacity to lead or direct others in the performance of their jobs.

In this regard, it appears significant that eight of the fifteen subjects answered question twenty-nine with a response greater than zero when the questionnaire was administered via distribution and without detailed verbal instructions. Additionally, fourteen of the fifteen subjects responded with a score greater than or equal to two on

question fifty-five when this same data collection method was employed. Other similar inconsistencies were evident, revealing that data obtained via distribution of the task analysis inventory could not be considered valid. Similar although less frequent inconsistencies appeared when the questionnaire was distributed to supervisors who employed it to report on the owrker characteristics required of their subordinates.

A possible explanation of these results might be attributed to the hesitancy of any one to admit that very few specific characteristics are required in the performance of his job. Quite naturally, humans associate with their jobs a complex set of procedures requiring a variety of human characteristics as well as a vast number of learned skills.

Although the number of subjects employed was small and not representative of the entire population, the personal interview appears to be the superior data collection method of those tested. Regardless of the explanations or conclusions involved, this portion of the paper represents an area of task analysis research that could and should be expanded.

IX. SUMMARY AND CONCLUSIONS

This paper has investigated some of the problems of present day training evaluation techniques, and has been devoted to describing methods for analyzing training effectiveness using valid task analysis inventory data. Hopefully, it may assist in bringing training emphasis more on target.

Task analysis data was collected by the investigator and subjected to parametric and nonparametric procedures for statistically determining the extent to which training was job oriented. Of particular interest was the matrix method of evaluation, which appears to be the most flexible procedure based on the nature of available data.

A general methodology for analyzing training programs using task analysis was presented. Included in this methodology was the importance of operational considerations such as the timeliness of training. In this regard, an evaluation approach originating with the classification of tasks by the matrix evaluation method was presented.

Although three specific supply oriented training programs were subjected to the analysis techniques presented, it was obviously impractical to assign an effective or ineffective rating to any one. A study of the data did, however, suggest the necessity for some changes in training emphasis provided the administrative and cost constraints

could be satisfied. Of particular interest was that the data revealed overtraining as significant a problem as undertraining. Although cost estimates were not obtained, this particular aspect of training analysis could and should be expanded as a possible approach toward significant training cost reductions.

Additionally three distinct questionnaire oriented data collection procedures were subjected to a small scale experiment to determine if any one method provided the most valid information. Although this experiment was based on a sample size of only fifteen relatively low ranking enlisted Marines, it was concluded that data can best be obtained by trained observers completing the proposed questionnaire based on verbal responses from the job incumbents. The possibility for future research in this area was also indicated.

Finally, this paper has attempted to indicate the importance of improved procedures for determining the adequacy of training content. Curriculum development would certainly profit from the use of task analysis data as feedback to determine whether training programs and objectives satisfactorily meet job specifications.

APPENDIX A

CURRICULUM DESCRIPTIONS AND ASSOCIATED RANK REQUIREMENTS

BASIC SUPPLY COURSE: This course is designed to train Privates (E-1) through Lance Corporals (E-3), who have limited or no supply experience, with only the basic fundamentals of the technical skills required of supply clerks. Basic skills taught include the preparation of routine correspondence and reports, use of supply publications and stock lists, computation of allowances, and a familiarization of basic property control and small purchase procedures.

SUPPLY ADMINISTRATIVE COURSE: This course is designed to train Corporals (E-4) through Staff Sergeants (E-6) with some of the formal principles, procedures and techniques related to the supply field. The course provides an intermediate level of instruction. Basic tasks taught include the following:

1. Additional emphasis on the tasks included in the Basic Course.

2. Supply operational, clerical, and administrative duties incident to requisitioning, procuring, and accounting for supplies and equipment.

3. The duties of a stock reviewer, maintaining card files, as well as screening and offsetting transaction cards with appropriate item record cards.

4. Mathematical operations on the electric calculator.

5. The general operating principles, uses, and capabilities of supporting EAM/EDP equipment.

6. Interpretation of computer produced reports.

7. Preparation of correspondence, messages, and reports in a format for future mechanization.

SENIOR COURSE: This course is designed to provide senior staff non-commisioned officers with a more complete understanding of the entire range of operational principles and methods which contribute to effective management within the following specific task elements

1. Supervising the duties of enlisted personnel.

2. Coordinating the daily supply activities of a unit or organization.

3. Supervising the duties of personnel involved in purchasing and contracting activities of a unit or organization.

4. Possessing a general knowledge of procedures, directives and regulations applicable to the functional areas of the supply system.

5. Assisting in the determination of supply and inventory management programs.

6. Possessing the required knowledge for the administration and operation of a supply activity or organization.

7. Having an understanding of basic computerized operations and their application to supply management techniques.

APPENDIX B

HUMAN FACTORS IN TASK ANALYSIS

A number of the human characteristics listed below may be needed in the satisfactory performance of the tasks included in your job. The accompanying form is designed to gauge your estimation of the necessity of the characteristics, which are referred to as Worker Characteristics.

In this section devoted to Worker Characteristics, indicate, by the placement of an "X" in the appropriate column, the degree of each characteristic you believe necessary to do your job. The definitions of the several degrees are defined as follows:

(0) This characteristic is not required.

(1) This characteristic is occasionally required.

(2) This characteristic is frequently required, but not on a regular repetitive basis.

(3) This characteristic is regularly required, the lack of which renders your job virtually impossible to perform.

Level Required
0 1 2 3

I. PHYSICAL

A. STRENGTH: In the four categories below, rate only muscular strength, not muscular endurance. Considerations: The size weight and bulkiness of objects handles; the pressure necessary to operate controls and tools.

1. Finger, hand, wrist and forearm strength -- That necessary to squeeze, bend, pull, twist, turn, shape or grip objects.
2. Upper arm strength -- That necessary to lift, swing, push, pull, carry or throw objects.
3. Back and shoulder strength -- That necessary to lift objects from the floor, move objects with the back and shoulders or swing heavy tools to strike objects.
4. Leg, foot and ankle strength -- That necessary to lift objects using knee action, operate pedals with pressure, grip or brace with the knees or climb, kneel, walk or stand with loads.

B. STAMINA: In the following six categories, consider both muscular endurance, the ability to sustain strength over a period of time, and circulo-respiratory endurance, that characteristic commonly referred to as wind, the ability to sustain vigorous activity over a period of time, including the maintenance of strength. Consideration: Pacing by machines or superiors; the frequency, duration and rapidity of movement; the extent of vigorous activity such as lifting, climbing, running, crawling, leaping, etc.;

Level Required
0 1 2 3

whether the duration of the activity is known or indeterminate length.

5. Rapid work for a series of short periods.
6. Rapid work for extended periods.
7. Rapid work for indefinite periods.
8. Heavy work for a series of short periods.
9. Heavy work for extended periods.
10. Heavy work for indefinite periods.

C. DEXTERITY: The three categories immediately below rate the skill or adroitness in the movement of the subject parts of the body. Considerations: the speed, complexity, and repetitiveness of the movements; the accuracy required; whether or not all digits or limbs are used to complete the movement.

11. Finger dexterity -- Ability to move and manipulate objects.
12. Hand and arm dexterity (including the fingers as part of the whole). Ability to move hands and arms accurately and quickly.
13. Foot and leg dexterity -- Ability to move feet and legs accurately and quickly.

D. COORDINATION: Considerations in the following five categories include the frequency, complexity and repetitiveness of movements.

14. Eye-hand coordination -- That necessary to control the hand through the use of vision.

Level Required
0 1 2 3

15. Foot-eye-hand coordination -- That necessary to control, simultaneously, the feet and hands through the use of vision.

16. Coordination of the independent movement of both hands -- That necessary to control, simultaneously and independently, both hands, without necessarily using the eyes. Additional considerations are the distance the hands move and the differences in the movement of the two hands. Each hand may be doing something different.

17. Foot-eye coordination -- That necessary to control the feet through the use of vision.

18. Foot-hand coordination -- That necessary to control the feet and hands simultaneously, not using the eyes.

E. SIZE CONSIDERATIONS: Any boundaries on your physical size, upper or lower, necessitated by the task performed, not by regulations.

19. Height. (fill in blank) _____.

20. Weight. (fill in blank) _____.

F. WORKING CONDITIONS:

21. Unpleasant conditions -- To work under conditions affecting physical comfort. These include bad odors, excessive noise, vibration, dust, dirt, fumes or moisture, high humidity, extremes in temperature, constant flux or a wide range in temperature, exposure to acids, unpleasant sights.

Level Required			
0	1	2	3

22. Hazardous conditions -- To work under conditions affecting physical safety.

II. SENSORY CHARACTERISTICS

A. VISUAL:

23. General visual acuity -- That necessary to perceive or recognize objects, locate points at a distance or make accurate discriminations using vision.

24. Color discrimination -- That necessary to distinguish similarities and differences in colors and shades thereof, create harmonious combinations of color or mix or match colors.

B. VISUAL - MENTAL: The five categories below involve judgments requiring the use of vision. Considerations: the complexity of the objects perceived; the frequency and rapidity of the required observations; their variability; any mechanical aids used.

25. Size estimation -- That necessary to make accurate judgments of dimensions (height, weight, depth, breadth or thickness) or over-all size, area.

26. Quantity estimation -- That necessary to make accurate judgments of the number or capacity of objects.

27. Speed estimation -- That necessary to make accurate judgments of rate of change of position, involving the estimation of both time and distance.

Level Required

0 1 2 3

28. Quality estimation -- That necessary to judge the quality or workmanship of material. This characteristic may very well involve the use of the other senses, such as touch, hearing, smell or taste.

29. Form perception -- That necessary to distinguish correct shape or outline or generally perceive shape.

C. AUDITORY

30. General keenness of hearing -- That necessary to recognize particular sounds and distinguish similarities and differences in pitch, intensity and quality of sounds.

D. OLFACTORY

31. Sense of smell -- That necessary to recognize particular odors or discriminate differences and similarities in the quality or intensity of odors.

E. TACTILE

32. Tough discrimination -- That necessary to accurately judge smoothness, roughness and other surface qualities, using touch.

33. Muscular discrimination -- That necessary to make judgments based on muscular sensitivity, such as estimating weight or resistance by lifting, pushing or pulling, gauging position or guiding body members without the use of vision.

Level Required
0 1 2 3

F. TASTE

34. Sense of taste -- That necessary to accurately distinguish similarities and differences in the intensity or quality of tastes, or recognize particular tastes.

III. MENTAL CHARACTERISTICS

A. MEMORY: The five characteristics below rate the capacity of the mind to store images for future references. Considerations: the number and complexity of images; their rate of acquisition; the length of time the images must be remembered.

35. Memory for concrete details.
36. Memory for ideas, theories, plans, processes, policies.
37. Memory for oral directions.
38. Memory for written directions.
39. Memory for names and persons.

B. LEARNED CHARACTERISTICS:

40. Arithmetic computation -- To perform calculations and higher mathematics.
41. Planning Ability -- To recognize and comprehend the steps necessary to achieve specific ends, decide upon, set up and coordinate such plans, and organize ideas and things.
42. Mechanical Ability -- To understand and use the principles of mechanical structure and operation and solve problems involving tools and machines.
43. Oral Expression -- To express oneself clearly and effectively, in use of speech.

Level Required
0 1 2 3

44. Written Expression -- To express oneself clearly and effectively, presenting information and ideas in writing.

C. PSYCHOLOGICAL CONSIDERATIONS

45. Attention to many items -- To keep in mind many parts of one job, although required to repeatedly and constantly shift one's attention from part to part.
46. Adaptability -- To adjust readily to new situations; flexibility.
47. Decision-making ability -- To be able to consider evidence, and, with a minimum of delay, reach a reasoned conclusion.
48. Initiative -- To recognize the need for a change in procedures and actions, and to accomplish such changes without specific instructions.
49. Tact -- To use diplomacy in dealing with people to achieve certain ends (consider only situations where this characteristic is fairly regular requirement of the job).
50. Personal appearance -- To maintain a level of neatness, grooming and attire above and beyond that considered necessary for a good Marine.
51. Concentration amid distractions -- To continue performing a task or tasks amid noise, interruptions or other distracting influences.

Level Required
0 1 2 3

52. Emotional stability -- To maintain self-control and calmness at all times.
53. Dealing with the public -- To meet and deal with the public, not necessarily involving the exercise of tact, but maintaining friendly relations.
54. Teamwork -- To necessarily subordinate one's individual performance to the good of the team or unit.
55. Leadership -- To cause subordinates to willingly produce desired results through a combination of superior knowledge, thoughtfulness, courage and exemplary personal performance.
56. Dependability -- To produce desired results at or prior to the time they are needed.
57. Physical courage -- To perform the necessary regardless of possible harmful physical consequences.
58. Moral courage -- To do what is right regardless of the consequences.
59. Please list below any other characteristics required in your job which have not been covered above. Then check the level of the characteristic required in your job.

APPENDIX C

PERFORMANCE TASK QUESTIONNAIRE

The section below consists of a list of actual tasks possibly included in your military occupational specialty. Place an "X" in the appropriate column to indicate the frequency with which you perform each task. The frequency is defined as follows:

(0) I do not perform this task.

(1) I perform this task occasionally.

(2) I perform this task frequently. It is a regular part of my job.

(3) I perform this task virtually to the exclusion of others included in my MOS.

TASK ELEMENTS:	0 1 2 3
A. PLANNING AND ORGANIZATION	
1. Organize and administer the supply activities for an FMF unit, post or station.	
2. Plan and coordinate the daily supply activities of FMF unit, post or station.	
3. Assign work to personnel on a daily basis.	
4. Direct the training of supply personnel.	
B. ADMINISTRATION	
5. Utilize the contents of basic directives and publications pertaining to supply.	
6. Utilize publications pertaining to subsistence accounting.	

7. Use the Marine Corps stack lists to verify stock numbers, nomenclature, units of issue and other data pertaining to supply.

C. APPLICATION

8. Prepare or edit requisitions, follow ups and cancellations for supplies and equipment.
9. Prepare or edit invoices and vouchers pertaining to property control records.
10. Monitor stock levels and post required entries to appropriate accounts.
11. Maintain financial records.
12. Assist in the preparation of a budget.
13. Reconcile fiscal listings and assist in the management of funds.
14. Conduct physical inventories of supplies and equipment.
15. Make adjustments to property records caused by investigations, inventories or other property accounting transactions.
16. Operate a typewriter.
17. Operate a desk calculator.
18. Establish and operate a routine filing system.
19. Maintain correspondence and directive files.
20. Employ the procedures for the open market purchase of supplies and equipment.
21. Maintain an allowance list for supplies and equipment.
22. Prepare routine correspondence and reports pertaining to supply.

23. Maintain informal account property records.
24. Compute reorder points and requisitioning objectives.
25. Establish or supervise a stock control system.
26. Screen and offset transaction cards with appropriate item record cards.
27. Know the general operating principles, uses, and general capabilities of supporting EAM/EDP equipment.
28. Interpret computer and EAM produced reports and take the required action.
29. Prepare correspondence or reports for key punching.
30. Employ the principles of field warehousing to include the proper care of material in storage.
31. Be familiar with the procedures for the shipment of Marine Corps material to include the services offered and the responsibilities of commercial carriers.
32. Know the principles and procedures for equipment maintenance utilized by the Marine Corps.

APPENDIX D

CUMULATIVE STEP FUNCTION DATA IN KOLMOGOROV-SMIRNOV TESTS

BASIC COURSE

Task No.	Training (CDF)	Job (CDF)	d_i
1	.02	.0018	.0182
2	.02	.0036	.0164
3	.02	.0250	.0050
4	.02	.0250	.0050
5	.09	.1015	.0115
6	.10	.1086	.0086
7	.28	.1816	.0984
8	.40	.2688	.1312
9	.51	.2955	.2145
10	.56	.3346	.2254*
11	.56	.3666	.1934
12	.56	.3755	.1845
13	.56	.3986	.1614
14	.56	.4591	.1009
15	.57	.5036	.0664
16	.60	.5890	.0110
17	.60	.6566	.0566
18	.60	.6993	.0993
19	.61	.7456	.1356
20	.67	.7634	.0934
21	.68	.7794	.0994
22	.73	.8257	.0957
23	.92	.8542	.0658
24	.96	.8791	.0809
25	.96	.8791	.0809
26	.96	.8846	.0754
27	.99	.9273	.0627
28	1.00	.9842	.0158
29	1.00	1.00	0
30	1.00	1.00	0
31	1.00	1.00	0
32	1.00	1.00	0

*D = .2254

ADMINISTRATIVE CURRICULUM

Task No.	Training (CDF)	Job (CDF)	d_i
1	.025	.0101	.0149
2	.025	.0149	.0101
3	.025	.0596	.0346
4	.025	.0719	.0469
5	.080	.0417	.0617
6	.095	.1453	.0503
7	.215	.2204	.0054
8	.290	.2830	.0070
9	.330	.3053	.0247
10	.350	.3482	.0018
11	.400	.3740	.0260
12	.435	.3863	.0487
13	.445	.4030	.0420
14	.455	.4210	.0340
15	.465	.4657	.0007
16	.485	.5319	.0469
17	.510	.5838	.0738
18	.515	.6160	.1010
19	.530	.6625	.1325
20	.560	.6726	.1126
21	.565	.7002	.1352
22	.595	.7628	.1678
23	.640	.7968	.1568
24	.645	.8226	.1776
25	.645	.8274	.1824*
26	.695	.8375	.1425
27	.790	.8751	.0851
28	.820	.9306	.1106
29	.840	.9529	.1129
30	.910	.9761	.0661
31	.970	1.00	.0300
32	1.00	1.00	0

*D = .1824

SENIOR COURSE

Task No.	Training (CDF)	Job (CDF)	d_i
1	.0294	.0351	.0057
2	.0412	.0652	.0240
3	.0586	.1364	.0778
4	.0645	.1886	.1241
5	.1294	.2538	.1244
6	.1294	.2628	.1334
7	.2533	.3230	.0697
8	.2827	.3672	.0845
9	.3063	.3902	.0839
10	.3122	.4062	.0940
11	.3712	.4213	.0501
12	.4125	.4473	.0348
13	.4243	.4774	.0531
14	.4243	.5075	.0832
15	.4479	.5325	.0846
16	.4479	.5646	.1167
17	.4479	.5997	.1518
18	.4479	.6379	.1900
19	.4479	.6790	.2311
20	.4713	.7111	.2398
21	.4772	.7412	.2640
22	.5480	.8014	.2534
23	.5598	.8274	.2676
24	.5598	.8334	.2736*
25	.6129	.8535	.2406
26	.6129	.8625	.2496
27	.8666	.9077	.0411
28	.9138	.9578	.0440
29	.9492	.9749	.0257
30	.9492	.9799	.0307
31	.9728	.9900	.0172
32	1.00	1.00	0.0

*D = .2736

APPENDIX E

TASK TRAINING AND JOB IMPORTANCE RANKINGS FOR COMPUTATION OF SPEARMAN RANK CORRELATION COEFFICIENTS

BASIC COURSE

Task No.	Training Rank	Job Rank
1	12	25.5
2	25	25.5
3	25	18
4	25	30
5	5	3
6	15	23
7	2	4
8	3	1
9	4	16
10	7.5	13
11	25	14
12	25	22
13	25	18
14	25	6
15	15	10
16	10.5	2
17	25	5
18	25	11.5
19	15	8.5
20	6	19.0
21	15	20.5
22	7.5	8.5
23	1	15
24	9	17
25	25	30
26	25	24
27	10.5	11.5
28	15	7
29	25	20.5
30	25	30
31	25	30
32	25	30

ADMINISTRATIVE COURSE

Task No.	Training Rank	Job Rank
1	16.5	27.0
2	30.5	29.5
3	30.5	9.5
4	30.5	24.5
5	6.0	2.0
6	23.5	31.0
7	1.0	1.0
8	3.0	4.5
9	10.5	19.5
10	19.0	11.0
11	7.5	16.5
12	12.0	24.5
13	23.5	23.0
14	23.5	22.0
15	23.5	9.5
16	19.0	3.0
17	16.5	7.0
18	27.0	14.0
19	21.0	8.0
20	14.0	27.0
21	27.0	15.0
22	14.0	4.5
23	9.0	13.0
24	27.0	16.5
25	30.5	29.5
26	7.5	27.0
27	2.0	12.0
28	14.0	6.0
29	19.0	19.5
30	4.0	18.0
31	5.0	21.0
32	10.5	32.0

SENIOR COURSE

Task No.	Training Rank	Job Rank
1	12.5	11.5
2	20.0	16.5
3	17.5	1.0
4	23.0	5.0
5	4.0	2.0
6	28.5	29.5
7	2.0	3.5
8	12.5	8.0
9	15.0	22.0
10	23.0	26.0
11	6.0	27.0
12	9.0	19.5
13	20.0	16.5
14	28.5	16.5
15	15.0	21.0
16	28.5	13.5
17	28.5	11.5
18	28.5	10.0
19	28.5	9.0
20	10.5	13.5
21	23.0	16.5
22	3.0	3.5
23	20.0	19.5
24	28.5	31.0
25	7.0	23.0
26	28.5	29.5
27	1.0	7.0
28	8.0	6.0
29	10.5	25.0
30	28.5	32.0
31	15.0	28.0
32	17.5	24.0

LIST OF REFERENCES

1. Bradley, James V., Distribution Free Statistical Tests, New Jersey: Prentice-Hall, Inc., 1968, pp. 15-44.

2. Freund, John E. and Miller, Irwin, Probability and Statistics for Engineers, New Jersey: Prentice-Hall Inc., 1965, pp. 162-165.

3. Naval Postgraduate School Research Report, Human Factors Data Collection Research Project, Professor G.K. Pook, October, 1969.

4. Seigel, Sidney, Nonparametric Statistic for the Behavioral Sciences, New York: McGraw Hill, Inc., 1956, pp. 47-51.

5. Ibid., pp. 202-213.

6. Ibid., pp. 68-75.

7. Seigel, Arthur I., Schultz, Douglas G., and Federman, Phillip, Training Performance Criterion Development and Application: A Matrix Method for the Evaluation of Training, Wayne, Pennsylvania: Applied Psychological Service, January, 1961.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Professor Gary K. Poock, Code 55 Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1
4.	Capt. Joseph D. Stewart 201 Old County Road Severna Park, Maryland 21146	1
5.	Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Naval Postgraduate School
Monterey, California 93940

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

THE USEFULNESS OF TASK ANALYSIS IN THE EVALUATION OF MILITARY TRAINING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Master's Thesis; September 1970

5. AUTHOR(S) (First name, middle initial, last name)

Joseph David Stewart, Captain, United States Marine Corps

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1970	84	7

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

Determining the proper emphasis of curriculum contents as well as judging the value or worth of training programs has become an important problem. The purpose of this paper is to demonstrate the usefulness of task analysis in measuring the effectiveness of training courses based on the extent to which curriculum contents are job oriented. In this regard, parametric and nonparametric statistical procedures are discussed as well as a matrix method of evaluation. A general methodology to include the operational significance of the data is also included.

Additionally, the results of a small scale experiment are presented. This experiment was conducted to determine the most valid questionnaire associated data collection method.

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
TASK ANALYSIS						
JOB ANALYSIS						
TRAINING EVALUATION						

120884

Thesis
S7146
c.1

Stewart
The usefulness of task analysis in the evaluation of military training.

thesS7146
The usefulness of task analysis in the e
3 2768 001 00971 5
DUDLEY KNOX LIBRARY